Machine Learning
Assigment 4

Submitted by:
Nishika Chopra   nc2259   N1859308
Karanpreet singh wadhwa   ksw352   N13337853

1. List the 5 tokens that occur most frequently in the training set.
Ans. [('.'), (','), ('the'), ('a'), ('and')]

2.Using just the training documents, calculate the information gain of every attribute. List the 5 attributes with the highest information gain.
Ans. Five attribute with highest gain are:
   ['bad', 'best', "n't", 'too', 'moving']

3. Using only the 50 attributes with highest information gain, train a neural net on the training examples. Your neural net should have one hidden layer with sigmoid units and a single output node. It should use the sigmoid function as the activation function in both the hidden units and the output unit, with cross-entropy as the loss function. You may modify the bppy.py code to do this, if you like. It is up to you to determine the correct learning rate and the number of hidden nodes. Do this using only on the training set. Then use the neural net with the learned weights (and the same set of attributes) to predict the outputs on the test set. Report your results on the test set using a confusion matrix. Also, list the percentage accuracy obtained.

ANS.
Atta used=0.03
No of feature=50
Hidden nodes=100
Epochs=1000

	Predict 0	Predict 1
Correct 0	127	100
Correct 1	64	209

Accuracy = 67.2%

4. What percentage accuracy is achieved if you use Zero-R, instead of a neural net?

Ans. Accuracy is 54.6 %

5. We used only the top 50 attributes, but we could have used the top k attributes, for larger or smaller k. Why is it reasonable to think that increasing the number of attributes might increase accuracy? Why is it reasonable to think that decreasing the number of attributes might increase accuracy?

Ans.
1. Increasing the number of attributes might increase accuracy because with it neural net would have more data to learn from an example and hence can predict more accurately.
2. Decreasing the number of attributes might also increase accuracy as it may prevent it from overfitting on training data. (some time there are feature which do nothing in differentiating between example like '.' Etc. removing feature like this would reduce the learning time also!)

6. Perform experiments to see how accuracy changes as you vary the number k of attributes for your neural net. Choose at least 4 values of k (in addition to k = 50) and graph the results. The horizontal axis should correspond to the number of attributes, and the vertical to the test accuracy. Did the results surprise you? Did you have any difficulties running the experiments? Give the graph AND the answers to these two questions.

Ans.

Number of feature vs accuracy
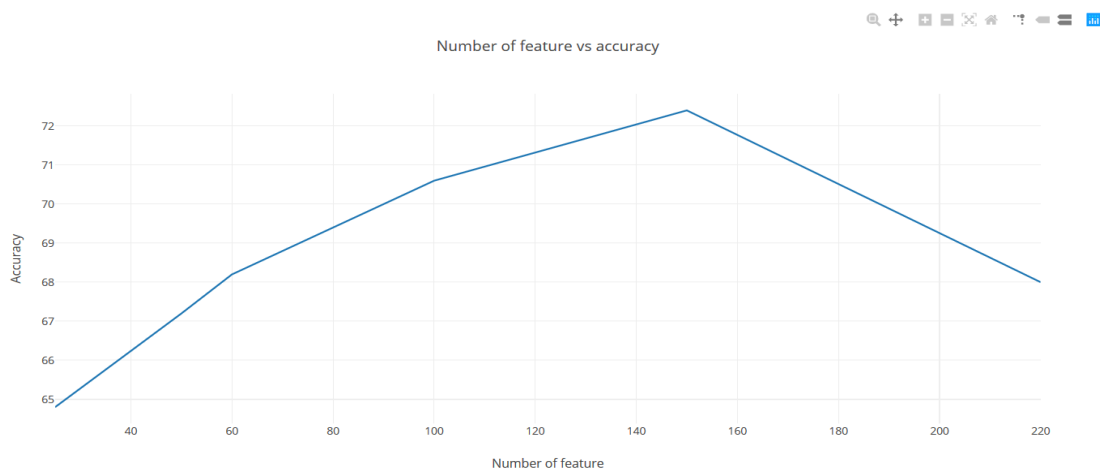
Number of Features	25	50	60	100	150	200
Accuracy	64.8	67.2	68.2	70.6	72.4	68.75

yes, the result surprised me. Initially, value of accuracy was increasing with Number of feature but after we keep on increasing the value of number of features accuracy started decreasing.

No, I didn't have difficulties while running the experminet.